THE NEW USE OF ADMINISTRATIVE DATA IN THE CENSUS OF MANUFACTURING BY STATISTICS NEW ZEALAND

A. G. Bullivant,
Statistics New Zealand, New Zealand

ABSTRACT

For the first time Statistics New Zealand will be combining data coming from the taxation administration system with data collected from a sample postal survey run by Statistics New Zealand (SNZ). This method is being used to produce census results for the manufacturing industry. Approximately three quarters of the population will not receive a SNZ questionnaire but instead the unit data for these manufacturers will be coming from the tax system data. The characteristics of these businesses are that they employ few people and are not owned by or own any other business. This results in a significant decrease in respondent burden with little degradation in coverage or data quality.

The use of the administrative data in this way has only been possible due to major changes in both the tax office that supplies the data and in Statistics New Zealand. This paper describes these changes and how they have allowed this new use. The major change instigated by SNZ was the complete matching of the SNZ business frame to the administrative units on the tax agency's database. This allows a link between the administrative data and any other survey data that uses the business frame as a source for its population. This paper describes the development process, the method adopted for integrating the tax data with data collected in a traditional postal survey and concludes with a brief discussion of other present and future uses which Statistics New Zealand has identified for data coming from this administrative source.

KEYWORDS

Administrative Data, Respondent Burden, Manufacturing

The New Use of Administrative Data in the Census of Manufacturing by Statistics New Zealand

A.G. Bullivant[1]

Introduction

Statistics New Zealand (SNZ), as with many other statistical agencies is attempting to balance compliance costs with a continuing need to collect a wide range of data to meet clients' needs. This has forced SNZ to look more closely at the data collected by other agencies in particular that collected by the tax agency of New Zealand (Inland Revenue Department - IRD).

All government agencies in New Zealand have experienced considerable change in the last five years. For their part, SNZ and the IRD have undergone considerable changes centred around a greater client focus, efficiency gains and computerisation.

The changed environment which has allowed this development is described initially. Following this is a description of; the development process, the method adopted for integrating the tax data with data collected in a traditional postal survey and concludes with a brief discussion of other present and future uses which SNZ has identified for data coming from this administrative source.

This paper looks at the benefits that have arisen due to the interaction of SNZ and the IRD in this changed environment. To do this the paper focuses mainly on the development of the most recent use of IRD data by SNZ for the Manufacturing Census 1995.

The use of IRD data in the Manufacturing Census is not the first use SNZ has made of administrative data , but this is the first time that data from an administrative source has been combined with data collected by SNZ to produce an integrated output.

Changed Environment

Major restructuring of the IRD has included a total redesign of their computer systems, centralisation of form processing and a greater emphasis on documentation of coding rules and other work practices. This has resulted in improved access to data which is easier to manipulate and of a consistently higher quality.

This has led to SNZ viewing a greater set of possibilities for the use of IRD data.

SNZ has been using data from IRD in a limited way for many years, mainly for frame maintenance. However, this early work was through the matching of business names. There was no link between the SNZ or IRD databases. This led to inefficient merging of the two databases with the matching exercise having to be repeated each time the databases needed to be linked.

---

[1] A.G. Bullivant. Survey Design and Development Division, Statistics New Zealand, Private Bag 92003, Auckland, New Zealand

In 1992 SNZ began the matching and maintaining of the IRD unique reference number on the SNZ business directory. This exercise was completed in 1994. The business directory (BD) now has the potential to act as a gateway between the SNZ surveys that use the BD as a frame and the wealth of tax agency data.

The first attempt to utilise this link, for other than frame maintenance, was undertaken with the Manufacturing Census 1995.

In the past SNZ had undertaken a census of all industries in one year but due to resource constraints it had been decided to cover the economy in three areas; Distribution, Manufacturing and the Rest of the Economy[2] in three consecutive years. In addition to the cycle of censuses, there is an annual sample survey across the whole of the economy known as the Annual Enterprise Survey (AES).

The Census for the Distribution sector of the economy was conducted in 1992 and experienced major difficulties relating to low response rates. The low response rate was partially attributed to the large number of respondents in the survey who had only very infrequent experience of completing SNZ questionnaires (approximately 70% of respondents would only receive a financial questionnaire once every 5 years).

The IRD collect general accounts information to assist with auditing of tax paying businesses. The accounts information collected was considered similar to that required to satisfy the SNZ census requirement.

Due to the similarities in data collected by the IRD, the response problems of the Distribution Census (which were considered likely to be repeated in the Manufacturing sector), and a general directive for government agencies to reduce respondent burden, it was decided to pursue the possibility of using tax data to replace surveying respondents in the Manufacturing Census.

Although there is now an established link to the IRD database there were a number of concerns relating to this new use of tax data which had to be investigated before the use could be put into practice.

Investigation of Quality Issues

The project team working on this development identified 22 quality issues relating to the use of IRD data in the Manufacturing Census. Some of the more significant issues were;

- The quality of data coming from the IRD
- The effect of method of supply on data consistency
- Mechanism to query respondents
- Satisfactory response rates

---

[2] Coverage Distribution sector covers wholesale trade, retail trade, restaurants and hotels, and Personal and Household Services
Manufacturing sector covers all manufacturers
Rest of Economy sector covers the remainder of the economy except Agriculture production, Central and Local Government, and religious organisations which are excluded.

- Consistency between IRD collection units and SNZ collection units

These significant quality issues are described in more detail below. At the end of this section there is a description of the method used for determining if it was acceptable to carry on with the development.

The quality of data coming from the IRD

The IRD form which provides the data for this study is used by the IRD in audit situations and for tax forecasting. It is not primarily used for tax collection. Because the form is of secondary use to the IRD, they do not have any edit checks on the data keyed. The only checks they perform are to ensure the information has been correctly transcribed from the form.

With no editing being performed the project team treated the data as raw. When edits were applied to the data, the proportions of units failing edits was similar to that expected if the data had been collected by SNZ.

Receiving the data in this raw state was seen as a bonus as there is no confounding effect of any alterations by IRD resulting from their own checks and SNZ could treat the data as though it had been received directly by SNZ from the respondent.

The effect of method of supply on data consistency

The IRD data used for this exercise enters the IRD database from three different sources; as an IRD form completed by an IRD employee from annual accounts sent by the respondent, or as an IRD form completed by the respondent and posted back, or as an IRD form completed by the respondent and electronically filed.

There were distinct differences between the data coming from different sources. Data that had been completed by IRD employees from annual accounts was subject to rigorous interpretation of the IRD coding instructions. Unfortunately the coding instructions were flawed in several areas and produced data that was not suitable for SNZ. This was pointed out to to the IRD and with some changes to coding notes the problem has been alleviated.

Forms completed by respondents and returned by mail had in some cases had data that was effected by adherence to the incorrect coding instructions. This usually occurred when the form was completed by a tax practitioner on the behalf of the taxpayer. Tax practitioners were also supplied with the coding instructions.

In data received from respondents that filed their tax returns electronically there was a high occurrence of null responses which were interpreted as zero values by the IRD processing system. Quite often the data sent electronically had been generated directly from an accounting software package and some of these were set up solely to complete the primary tax collection form and not the accounts information form that supplied the data of interest to SNZ. It was therefore necessary to treat full null or zero responses as non-responses rather than valid responses.

Mechanism to query respondents

As in other cases where administrative data is used by the statistical agency, it was not possible

for SNZ to directly query respondents supplying data to IRD. It was therefore necessary to develop alternative mechanisms.

If a respondents data failed most input edits applied by SNZ, the unit was treated as a non-response. If the unit had a small number of edit failures and it was clear that a simple error had been made i.e. addition etc., then the data would be adjusted.

Satisfactory response rates

Primary tax collection forms used by the IRD are compulsory to complete and have high response rates. The data used in this study is collected by the IRD using a secondary tax collection form which is not compulsory.

From an analysis of the data used for this study it was expected that there would be approximately a 75% response rate. This compares favourably with the 68% response of the Distribution Census run in 1992. The expected response rate for the IRD data was therefore, considered acceptable.

Consistency between IRD collection units and SNZ collection units

The exercise of matching IRD reference numbers to the SNZ business directory indicated that the IRD collection unit was analogous in most cases to the enterprise unit on the business directory. In SNZ's financial collections the collection unit is the accounting unit. In most cases the accounting unit is the same as the enterprise, but for some enterprises there is more than one accounting unit. In these cases it is common for the accounting units to be classified to different industries.

There was some consideration put to the idea of estimating the proportion of the tax data going to each of the accounting units. This apportionment would have used employment as a basis but this was considered to be an overly complex process which had the potential to produce spurious data. Therefore it was considered inappropriate to use tax data to replace data collected by SNZ for those enterprises with more than one accounting unit.

Tax data for individual units was compared to data for the same units coming from the AES survey. This comparison highlighted that in some instances the responses being filed with the IRD were not for the enterprise that the IRD unit had been matched to on the business directory. These responses were in fact for a group of enterprises linked by common ownership. In these cases the IRD unit corresponded to the SNZ Group Top Enterprises (GTE).

A GTE is a group of enterprises which are linked by ownership links of 50% or more. As with the multi-accounting enterprises, it was common for the enterprises that made up the GTE to be from different industries. Again it was decided not to use tax data for these types of units.

Determining whether to proceed

After working through the above quality issues it became apparent that the best way of proving the ability of the IRD data to perform as expected was to trial the production of results using the data.

The trial was performed on three sub industries within Manufacturing. It mimicked as closely as possible the proposed method of using earlier period IRD data. The resulting data from the trial was then evaluated.

There were two methods of appraisal;

1. The data was used to provide census information for a trial sample re-selection of a quarterly Manufacturing Survey run by SNZ. The results of the sample re-selection using the tax data were similar to those obtained from a re-selection performed earlier using SNZ census information.
2. Comparison of summarised totals produced using IRD data with results produced from the AES sample survey. For most variables the IRD data based totals were within the 95% confidence limits for the AES based results. For the remaining variables the cause of the disparity was linked to the errors in the coding files discussed in the quality issue above.

Statistics New Zealand was therefore confident that the taxation data was of satisfactory quality to be used for the 1995 Manufacturing census. The next section reviews the implementation of the mix of taxation and survey data.

Implementation Method Used and Issues

The implementation method covers the population of the Manufacturing Industry in four subsets. It was decided to use data from three sources to cover the population; from AES, from additional questionnaires sent by SNZ, and from IRD data.

Diagram 1 - Breakdown of the Manufacturing Census Population showing the areas where different collection methods were used.

Manufacturing Census Population

THE REST (IRD Data)

There were three types of units in the population for which it was considered inappropriate to use IRD data to replace data collected by SNZ. These were enterprises that had multiple accounting units, enterprises that were owned by or owned other enterprises (GTEs) and enterprises that are not required to complete the IRD form (these were non-profit organisations and were few in number).

The AES sample covers all large businesses, all multiple accounting unit businesses, and samples of medium and small businesses. The remainder of the population was covered using IRD data except for units identified in subset 2 and 3 in the above diagram. These units were sent a SNZ questionnaire to obtain the information required.

The proportion of the population covered by each of the different subsets is shown in the table below.

Table 1 Breakdown of respondent in each subset

Subset	Approximate numbers in the population	Percentage of population covered
IRD data to be used	15,500	72%
1. AES sample	5,000	25%
2 Enterprise belongs to a GTE	500	3%
3. Enterprise not completing IRD form	30	0.1%

NB The total number of enterprises in the New Zealand economy is approximately 340,000.

The resulting data for the Manufacturing Census will be a mixture of taxation data and survey data. As the AES survey collects much more detail than is available from the IRD data, information from the survey will be summarised to the same level of detail as available from the IRD data.

At the time of writing this paper approximately 16% of the tax data respondents had been loaded to the SNZ database. These units are undergoing unit level quality checks using standard edit parameters. The results from the early stages of the practical implementation of this method suggest that the method is feasible in practice.

Other Potential Uses for Taxation Data

There have been many uses identified for the wide range of data available from the IRD.

At present SNZ is developing the use of Goods and Service Tax (GST) data. GST is a uniform tax applied to all sales of goods and charges for services. It is collected by businesses at the point of transaction and then paid by almost every business in the economy to the IRD. The aim of this project is to develop a publishable series using the sub-annual sales information available for the GST collection forms. The GST data would be linked with SNZ industry and region codes for each unit.

It is compulsory for businesses to register for GST if they have sales greater than $NZ 30,000 per annum. The information on GST registrations of new businesses is used to update the SNZ Business Directory. In addition, the value of sales reported on the GST return is used to determine the economic significance of businesses on the business directory.

In another project SNZ is reviewing the employment related business statistics it collects. It is clear that information available from the IRD Pay As You Earn (PAYE) tax data collected form businesses is going to provide an integral part of this development. PAYE tax is collected by every employing business from their employees salaries.

There are likely to be many other uses of tax data developed by SNZ and other government agencies in the near future.

Conclusion

The method of using tax data in the manufacturing census is being implemented at the moment. The true measure of success of this use of IRD data will be known at the completion of the Manufacturing Census 1995 when users start to analyse the data. However the indications are clear that the use of taxation data, will provide a method of reducing compliance cost, while at the same time still producing good quality business statistics.

These and future developments have only been possible due to these changes in the business environment in which SNZ collects statistics.

- The full matching of the business directory to the IRD reference number
- The ability to link the IRD data with SNZ survey data using the business directory as the link. This has been possible due to the advances in computer systems in both agencies.
- The strong working relationship between SNZ and IRD. This relationship is assisted by the strong directive of government to increase efficiency in government agencies and to reduce compliance costs to businesses.

SNZ use of the data is seen by IRD as very much a subsidiary use of the data compared to the main purpose of assisting in tax collection. Complete reliance on this data source for statistical uses is a risk, as the data could be effected by directives coming from the IRD with little input from SNZ. SNZ has recognised this risk and has worked to minimise it. This is been done through developing joint consultative committee between the two agencies and through promoting written contracts between the two agencies for data supply.

While SNZ has the advantage of fewer units to collect from than most statistical agencies, the use of data coming from IRD is seen by SNZ as the future direction for many of its present collections and is seen as having great potential for the production of new statistics.